

# OPEN SOURCE IS AT THE CORE OF MODERN SOFTWARE

Innovation through open collaboration has changed the technology industry forever

## OUR COMMITMENT

At Continuum, we value open source software and believe it is a privilege to be able to share idea-as-code with people around the world as we work together to build useful tools and products. We believe in building on the shoulders of giants and seek productive, sustainable ways to continue to strengthen the open source foundation and create the architecture of the future.

Travis Oliphant, our co-founder and CEO, is a key figure in the development of NumPy and SciPy. Most of our developers have a long history as open source contributors and have spent many, many volunteer hours writing open source software, giving talks at conferences, writing documentation, fixing bugs, answering questions, sharing thoughts on public mailing lists, and generally working to ensure that ideas-as-code get shared as far and widely as possible. Our team of contributors incubate Continuum Analytics projects and contribute to OSS projects.

We take great satisfaction in empowering people to solve some of the world's greatest challenges to make the world a better place and improve the lives of others.

## INCUBATING NEW PROJECTS TO MEET ENTERPRISE NEEDS

While much of the software we write at Continuum Analytics is open source from the beginning, some of our software is not freely available at first. Our sincere desire is for the features in our proprietary software to end up as open source software as soon as time and resources allow us to make it freely available. We believe that as a company, we can contribute to open source software best by providing livelihoods for developers that allow them to focus on writing software that gets contributed to open source in time. Sometimes, this means keeping software proprietary so that users of our software are also customers driving its development. This enables us to employ people full-time on the creation and support of software that effectively becomes open source as more people buy it. This gives more and more people a chance to participate in the creation of open source software — not just developers with spare cycles.

If you find our software useful, we hope that you will download it and be satisfied not only with the software itself, but also with the knowledge that you are contributing to the present and future ecosystem of open source software.



Blaze

**Blaze**  
A framework for automatic distribution and parallelization of Python



**Bokeh**  
A framework for plots, interactive and real-time streaming visualizations

conda

**conda**  
Package management for Python, R, NumPy, and SciPy

Dask

**Dask**  
A framework that enables parallelization of algorithms on modern architectures



**DyND**  
A library for dynamic in-memory arrays that extends the NumPy data model

numba

**numba**  
Compiles NumPy and SciPy into machine code



**PhosphorJS**  
A framework for building high performance, pluggable, desktop-style web applications

## BLAZE SCALES PYTHON ANALYTICS TO BIG DATA ON MULTIPLE COMPUTE ENGINES

Fast, scalable out-of-core computations on Big Data that extends NumPy and Pandas to distributed and streaming datasets

Blaze extends NumPy's successful model of array-oriented programming to out-of-core and distributed data. Blaze allows analysts and scientists to productively write robust and efficient code, without getting bogged down in the details of how to distribute computation, or worse, how to transport and convert data between databases, formats, proprietary data warehouses, and other silos.

The core of Blaze is a generic N-dimensional array/table object with a very general "data type" and "data shape" descriptor for all kinds of data, but especially semi-structured, sparse, and columnar data. Blaze's generalized calculation engine can iterate over the distributed array or table and dispatch to low-level kernels, selected via the dynamic data typing mechanism.

Blaze supports data stores and stream engines including:

- › Bcolz - compressed columnar
- › MongoDB – NoSQL
- › SQLAlchemy – Most SQL databases
- › Apache Spark – cluster computing framework
- › PyTables - high performance HDF5
- › Streaming Python - streaming data

**GET STARTED:** `conda install blaze`

**READ THE DOCS:** [blaze.pydata.org](http://blaze.pydata.org)

**GET SOURCE CODE:** [github.com/blaze](https://github.com/blaze)

## BOKEH SCALES VISUALIZATION TO BIG DATA

Interactive and real-time streaming visualization framework that scales to Big Data with data shading

Bokeh is a framework for creating versatile, interactive, browser-based visualizations of streaming or Big Data from Python, R, Scala, without writing any JavaScript. Its primary output backend is HTML5 Canvas.

There are many excellent plotting packages for Python, but they generally do not optimize for the particular needs of statistical plotting or multidimensional datasets. Additionally, advanced visual customization is typically difficult for non-programmers, and most libraries do not build a reliable data processing pipeline that supports rich interactivity like linked brushing. Bokeh addresses these problems at their core by using a declarative data transformation scheme and is engineered to operate in a client/server model for the modern web.

**GET STARTED:** `conda install bokeh`

**READ THE DOCS:** [bokeh.pydata.org](http://bokeh.pydata.org)

**GET SOURCE CODE:** [github.com/bokeh](https://github.com/bokeh)

## CONDA PACKAGES PYTHON, R, NUMPY, SCIPY EASILY

Eliminates package dependency and version control issues

Conda is an innovative package management tool that allows users to mix-and-match different versions of Python, NumPy, SciPy, and other packages in isolated environments and easily switch between the environments.

The conda command is the primary interface for managing Anaconda installations. It is great for solving enterprise integration and application deployment challenges. It can query and search the Anaconda package index and current Anaconda installation, create new Anaconda environments, and install and update packages into existing Anaconda environments.

**GET STARTED:** [continuum.io/tryAnaconda](https://continuum.io/tryAnaconda)  
**READ THE DOCS:** [conda.pydata.org](https://conda.pydata.org)  
**GET SOURCE CODE:** [github.com/conda](https://github.com/conda)

## DASK PARALLELIZES ANALYTICS ON MODERN MULTI-CORE MACHINES AND DISTRIBUTED CLUSTERS

Makes it easy to write complex parallel algorithms for task execution

Dask is a framework used to easily parallelize algorithms that takes advantage of the available memory and computer power to maximize memory, execution time, and performance of complex algorithms. Dask creates a task graph based on the data and then intelligently schedules the execution of the tasks to optimize throughput. While developers can parallelize Python manually, Dask helps to automate the task with rich primitives that are aware of the execution environment and optimize the analytic execution.

**GET STARTED:** `conda install dask`  
**READ THE DOCS:** [dask.pydata.org](https://dask.pydata.org)  
**GET SOURCE CODE:** [github.com/blaze/dask](https://github.com/blaze/dask)

## DYND IS DYNAMIC ETL-ON-READ FOR UNSTRUCTURED DATA

Make unstructured and semi-structured data just as easy to work with as structured data

DyND is a dynamic in-memory vector and array library for C++ and Python that makes it as efficient to process unstructured and semi-structured data as regularized structured data. This allows dynamic ETL-on-read to be performed where transformation operations are executed on CSV files, JSON, and any other unstructured or semi-structured data.

**GET STARTED:** `conda install blaze`  
**READ THE DOCS:** [continuum.io/os/dynd](https://continuum.io/os/dynd)  
**GET SOURCE CODE:** [github.com/libdynd](https://github.com/libdynd)

## NUMBA SPEEDS UP NUMPY AND SCIPY

Compiles Python into machine code for lightning fast execution

Numba is a compiled versions of NumPy and SciPy. It uses the LLVM compiler infrastructure to compile Python byte-code to machine code for use in the NumPy run-time and SciPy modules.

**GET STARTED:** `conda install numba`  
**READ THE DOCS:** [numba.pydata.org](https://numba.pydata.org)  
**GET SOURCE CODE:** [github.com/numba](https://github.com/numba)

## PHOSPHORJS SIMPLIFIES AND SPEEDS UP WEB APPS

A fast, flexible, and efficient web framework

PhosphorJS is a framework for building high performance, pluggable, desktop style web applications that integrates easily with existing web frameworks. The PhosphorJS framework has well-defined, efficient widgets and layouts that allow a developer to design high performance, responsive desktop style apps for the web that consistently achieve sub-millisecond layouts. This efficient design maximizes the execution time of business logic.

**READ THE DOCS:** [phosphorjspydata.org](http://phosphorjspydata.org)  
**GET SOURCE CODE:** [continuum.io/os/phosphorjs](http://continuum.io/os/phosphorjs)

## CONTRIBUTING TO OSS PROJECTS

Our developers enhance the Python ecosystem by actively contributing to a wide variety of open source software projects.



### Jupyter/iPython

Web application that allows you to create and share documents that contain code, visualization, and documentation



### NumPy

The fundamental package for scientific computing in Python



### Pandas

Package for high performance data analysis workflow in Python



### PyParallel

A Windows based multi-threaded CPython interpreter



### SciPy

A numerical integration and optimization package for Python



### Spyder

An integrated development environment (IDE) for Python



### SymPy

A Python package for symbolic mathematics



## SNAPSHOT OF OUR CONTRIBUTORS



# CONTINUUM<sup>®</sup> ANALYTICS

## ABOUT CONTINUUM ANALYTICS

Continuum Analytics develops Anaconda, the leading modern open source analytics platform powered by Python. Built on proven open source technology and easily integrated within existing IT environments, Anaconda allows organizations to make critical business decisions based on their data quickly, easily, inexpensively,

and with flexibility. Continuum's founders and developers have created or contribute to some of the most popular data science technologies, including NumPy, SciPy, Pandas, Jupyter/IPython, and many others. To learn more about the Anaconda platform, training and consulting services, visit [continuum.io](http://continuum.io)